

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: PELLETIER, Marc
BARKER, William A.
HAKES, David J.
ZOPF, David A.
- (ii) TITLE OF THE INVENTION: METHODS FOR PRODUCING
SIALYLLOLIGOSACCHARIDES IN A DAIRY SOURCE
- (iii) NUMBER OF SEQUENCES: 10
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: PENNIE & EDMONDS LLP
(B) STREET: 1155 Avenue of the Americas
(C) CITY: New York
(D) STATE: NY
(E) COUNTRY: USA
(F) ZIP: 10036-2711
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 2.0
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 08/911,393
(B) FILING DATE: 14-AUG-1997
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Coruzzi, Laura A
(B) REGISTRATION NUMBER: 30,742
(C) REFERENCE/DOCKET NUMBER: 7188-032-999
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (212)7909090
(B) TELEFAX: (212)8699741
(C) TELEX: 66141 PENNIE

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3183 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGGGAAAA	CAGTCGTTGG	GGCCAGTAGG	ATGTTCTGGC	TAATGTTTTT	CGTGCCGCTT	60
CTTCTTGCGC	TCTGCCCCAG	CGAGCCCGCG	CATGCCCTGG	CACCCGGATC	GAGCCGAGTT	120
GAGCTGTTTA	AGCGGCAAAG	CTCGAAGGTG	CCATTGAAA	AGGGCGGCAA	AGTCACCGAG	180
CGGGTTGTCC	ACTCGTTCCG	CCTCCCGGCC	CTTGTTAATG	TGGACGGGGT	GATGGTTGCC	240
ATCGCGGACG	CTCGCTACGA	AACATCCAAT	GACAACCTCC	TCATTGATAC	GGTGGCGAAG	300
TACAGCGTGG	ACGATGGGGA	GACGTGGGAG	ACCCAAATTG	CCATCAAGAA	CAGTCGTGCA	360

TCGTCTGTTT	CTCGTGTGGT	GGATCCCACA	GTGATTGTGA	AGGGCAACAA	GCTTTACGTC	420
CTGGTTGGAA	GCTACAACAG	TTCGAGGAGC	TACTGGACGT	CGCATGGTGA	TGCGAGAGAC	480
TGGGATATTC	TGCTTGCCGT	TGGTGAGGTC	ACGAAGTCCA	CTGCGGGCGG	CAAGATAACT	540
GCGAGTATCA	AATGGGGGAG	CCCCGTGTCA	CTGAAGGAAT	TTTTCCCGGC	GGAAATGGAA	600
GGAATGCACA	CAAATCAATT	TCTTGGCGGT	GCAGGTGTTG	CCATTGTGGC	GTCCAACGGG	660
AATCTTGTGT	ACCCTGTGCA	GGTTACGAAC	AAAAAGAAGC	AAGTTTTTTC	CAAGATCTTC	720
TACTCGGAAG	ACGAGGGCAA	GACGTGGAAG	TTTGGGGAGG	GTAGGAGTGA	TTTTGGCTGC	780
TCTGAACCTG	TGGCCCTTGA	GTGGGAGGGG	AAGCTCATCA	TAAACACTCG	AGTTGACTAT	840
CGCCGCCGTC	TGGTGTAAGA	GTCCAGTGAC	ATGGGGGAAT	CGTGGGTGGA	GGCTGTTCGG	900
ACGCTCTCAC	GTGTGTGGGG	CCCCTCACCA	AAATCGAACC	AGCCCGGCAG	TCAGAGCAGC	960
TTCACCTGCG	TGACCATCGA	GGGAATGCGT	CTTATGCTCT	TCACACACCC	GCTGAATTTT	1020
AAGGGAAGGT	GGCTGCGCGA	CCGACTGAAC	CTTGGCTGTA	CGGATAACCA	GCGCATTTAT	1080
AACGTTGGGC	AAGTATCCAT	TGGTGATGAA	AATTCCGCCT	ACAGCTCCGT	CCTGTACAAG	1140
GATGATAAGC	TGTACTGTTT	GCATGAGATC	AACAGTAACG	AGGTGTACAG	CCTTGTTTTT	1200
GCGCGCCTGG	TTGGCGAGCT	ACGGATCATT	AAATCAGTGC	TGCAGTCCTG	GAAGAATTGG	1260
GACAGCCACC	TGTCCAGCAT	TTGCACCCCT	GCTGATCCAG	CCGCTTCGTC	GTCAAGCCGT	1320
GGTTGTGGTC	CCGCTGTAC	CACGGTTGGT	CTTGTGTGGT	TTTTGTTCGA	CAGTGCCACC	1380
AAAACCGAAT	GGGAGGATGC	GTACCGCTGC	GTCAACGCAA	GCACGGCAAA	TGCGGAGAGG	1440
GTTCCGAACG	GTTTGAAGTT	TGCGGGGGTT	GGCGGAGGGG	CGCTTTGGCC	GGTGAGCCAG	1500
CAGGGGCAGA	ATCAACGGTA	TCACTTTGCA	AACCACGCGT	TCACGCTGGT	GGCGTCCGTG	1560
ACGATTACAG	AGGTTCCGAG	CGTCGCGAGT	CCTTTGCTGG	GTGCGAGCCT	GGACTCTTCT	1620
GGTGGCAAAA	AACTCCTGGG	GCTCTCGTAC	GACGAGAAGC	ACCAAGTGGC	GCCAATATAC	1680
GGATCAACGC	CGGTGACGCC	GACCGGATCG	TGGGAGATGG	GTAAGAGGTA	CCACGTGGTT	1740
CTTACGATGG	CGAATAAAAT	TGGTTCCGGT	TACATTGATG	GAGAACCTCT	GGAGGGTTCA	1800
GGGCAGACCG	TTGTGCCAGA	CGGGAGGACG	CCTGACATCT	CCCCTTTCTA	CGTTGGCGGG	1860
TATGGAAGGA	GTGATATGCC	AACCATAGAG	CACGTACCGG	TGAATAATGT	TCTTCTTTAC	1920
AACCGTCAGC	TGAATGCCGA	GGAGATCAGG	ACCTTGTTC	TGAGCCAGGA	CCTGATTGGC	1980
ACGGAAGCAC	ACATGGGCAG	CAGCAGCGGC	AGCAGTGCCC	ACAGTACGCC	CTCAACTCCC	2040
GCTGACAAAC	GTGCCCACAG	TACGCCCTCA	ACTCCCGCTG	ACAGCAGTGC	CCACAGTACG	2100
CCCTCAACTC	CCGCTGACAG	CAGTGCCCA	AGTACGCCCT	CAGCTCCCGG	TGACAACGGT	2160
GCCACAGTA	CGCCCTCAGC	TCCCCTGAC	AGCAGTGCCC	ACAGTACGCC	CTCAACTCCC	2220
GCTGACAAAC	GTGCCCACAG	TACGCCCTCA	GCTCCCGCTG	ACAGCAATGC	CCACAGTACG	2280
CCCTCGACTC	CCGCTGACAA	CGGTGCCCCAC	AGTACGCCCT	CAACTCCCGC	TGACAACGGT	2340
GCCCACAGTA	CGCCCTCAGC	TCCCCTGAC	AACGGTGCCC	ACAGTACGCC	CTCAGTCCC	2400
GGTGACAGCA	GTGCCCACAG	TACGCCCTCA	ACTCCCGCTG	ACAACGGTGC	CCACAGTACG	2460
CCCTCAGCTC	CCGCTGACAG	CAATGCCAC	AGTACGCCCT	CGACTCCCGG	TGACAACGGT	2520
GCCCACAGTA	CGCCCTCAGC	TCCCCTGAC	AGCAATGCC	ACAGTACGCC	CTCAGTCCC	2580
GCTGACAGCA	GTGCCCACAG	TACGCCCTCA	GCTCCCGCTG	ACAACGGTGC	CCACAGTACG	2640
CCCTCAGCTC	CCGCTGACAG	CAGTGCCCA	AGTACGCCCT	CAGCTCCCGG	TGACAACGGT	2700
GCCCACAGTA	CGCCCTCAGC	TCCCCTGAC	AACGGTGCCC	ACAGTACGCC	CTCAGTCCC	2760
GGTGACAGCA	ATGCCCACAG	TACGCCCTCG	ACTCCCGCTG	ACAGCAGTGC	CCACAGTACG	2820
CCCTCAACTC	CCGCTGACAG	CAGTGCCCA	AGTACGCCCT	CAGCTCCCGG	TGACAACGGT	2880
GCCCACAGTA	CGCCCTCAGC	TCCCCTGAC	AGCAGTGCCC	ACAGTACGCC	CTCAATTCCC	2940
GGTGACAGCA	GTGCCCACAG	TACGCCCTCA	GCTCCCGCTG	ACAACGGTGC	CCACAGTACG	3000
CCCTCAGCTC	CCGCTGACAG	CAGTGCCCA	AGTACGCCCT	CGACTCCCGC	TGACAACGGC	3060
GCTAATGGTA	CGGTTTTGAT	TTTGACAGAT	GGCGCTGCAT	TTTCGGCCTT	TTGGGGCGGA	3120
GGGCTTCTTT	TGTGTGCGGG	TGCTTTGCTG	CTGCACGTGT	TCGTTATGGC	AGTTTTTTTC	3180
TGA						3183

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1060 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Gly	Lys	Thr	Val	Val	Gly	Ala	Ser	Arg	Met	Phe	Trp	Leu	Met	Phe
1				5				10						15	
Phe	Val	Pro	Leu	Leu	Leu	Ala	Leu	Cys	Pro	Ser	Glu	Pro	Ala	His	Ala
			20					25					30		
Leu	Ala	Pro	Gly	Ser	Ser	Arg	Val	Glu	Leu	Phe	Lys	Arg	Gln	Ser	Ser
		35					40					45			

Lys	Val	Pro	Phe	Glu	Lys	Gly	Gly	Lys	Val	Thr	Glu	Arg	Val	Val	His
50						55					60				
Ser	Phe	Arg	Leu	Pro	Ala	Leu	Val	Asn	Val	Asp	Gly	Val	Met	Val	Ala
65					70					75					80
Ile	Ala	Asp	Ala	Arg	Tyr	Glu	Thr	Ser	Asn	Asp	Asn	Ser	Leu	Ile	Asp
				85					90					95	
Thr	Val	Ala	Lys	Tyr	Ser	Val	Asp	Asp	Gly	Glu	Thr	Trp	Glu	Thr	Gln
			100					105					110		
Ile	Ala	Ile	Lys	Asn	Ser	Arg	Ala	Ser	Ser	Val	Ser	Arg	Val	Val	Asp
		115					120					125			
Pro	Thr	Val	Ile	Val	Lys	Gly	Asn	Lys	Leu	Tyr	Val	Leu	Val	Gly	Ser
		130				135					140				
Tyr	Asn	Ser	Ser	Arg	Ser	Tyr	Trp	Thr	Ser	His	Gly	Asp	Ala	Arg	Asp
145					150					155					160
Trp	Asp	Ile	Leu	Leu	Ala	Val	Gly	Glu	Val	Thr	Lys	Ser	Thr	Ala	Gly
				165					170					175	
Gly	Lys	Ile	Thr	Ala	Ser	Ile	Lys	Trp	Gly	Ser	Pro	Val	Ser	Leu	Lys
			180					185					190		
Glu	Phe	Phe	Pro	Ala	Glu	Met	Glu	Gly	Met	His	Thr	Asn	Gln	Phe	Leu
		195					200					205			
Gly	Gly	Ala	Gly	Val	Ala	Ile	Val	Ala	Ser	Asn	Gly	Asn	Leu	Val	Tyr
		210				215					220				
Pro	Val	Gln	Val	Thr	Asn	Lys	Lys	Lys	Gln	Val	Phe	Ser	Lys	Ile	Phe
225					230					235					240
Tyr	Ser	Glu	Asp	Glu	Gly	Lys	Thr	Trp	Lys	Phe	Gly	Glu	Gly	Arg	Ser
				245					250					255	
Asp	Phe	Gly	Cys	Ser	Glu	Pro	Val	Ala	Leu	Glu	Trp	Glu	Gly	Lys	Leu
			260					265					270		
Ile	Ile	Asn	Thr	Arg	Val	Asp	Tyr	Arg	Arg	Arg	Leu	Val	Tyr	Glu	Ser
		275					280					285			
Ser	Asp	Met	Gly	Asn	Ser	Trp	Val	Glu	Ala	Val	Gly	Thr	Leu	Ser	Arg
		290				295					300				
Val	Trp	Gly	Pro	Ser	Pro	Lys	Ser	Asn	Gln	Pro	Gly	Ser	Gln	Ser	Ser
305					310					315					320
Phe	Thr	Ala	Val	Thr	Ile	Glu	Gly	Met	Arg	Val	Met	Leu	Phe	Thr	His
				325					330					335	
Pro	Leu	Asn	Phe	Lys	Gly	Arg	Trp	Leu	Arg	Asp	Arg	Leu	Asn	Leu	Trp
			340					345					350		
Leu	Thr	Asp	Asn	Gln	Arg	Ile	Tyr	Asn	Val	Gly	Gln	Val	Ser	Ile	Gly
			355				360					365			
Asp	Glu	Asn	Ser	Ala	Tyr	Ser	Ser	Val	Leu	Tyr	Lys	Asp	Asp	Lys	Leu
					375						380				
Tyr	Cys	Leu	His	Glu	Ile	Asn	Ser	Asn	Glu	Val	Tyr	Ser	Leu	Val	Phe
385					390					395					400
Ala	Arg	Leu	Val	Gly	Glu	Leu	Arg	Ile	Ile	Lys	Ser	Val	Leu	Gln	Ser
				405					410					415	
Trp	Lys	Asn	Trp	Asp	Ser	His	Leu	Ser	Ser	Ile	Cys	Thr	Pro	Ala	Asp
			420					425					430		
Pro	Ala	Ala	Ser	Ser	Ser	Glu	Arg	Gly	Cys	Gly	Pro	Ala	Val	Thr	Thr
		435					440					445			
Val	Gly	Leu	Val	Gly	Phe	Leu	Ser	His	Ser	Ala	Thr	Lys	Thr	Glu	Trp
		450				455					460				
Glu	Asp	Ala	Tyr	Arg	Cys	Val	Asn	Ala	Ser	Thr	Ala	Asn	Ala	Glu	Arg
465					470					475					480
Val	Pro	Asn	Gly	Leu	Lys	Phe	Ala	Gly	Val	Gly	Gly	Gly	Ala	Leu	Trp
				485					490					495	
Pro	Val	Ser	Gln	Gly	Gln	Asn	Gln	Arg	Tyr	His	Phe	Ala	Asn	His	
			500				505					510			
Ala	Phe	Thr	Leu	Val	Ala	Ser	Val	Thr	Ile	His	Glu	Val	Pro	Ser	Val
			515				520					525			
Ala	Ser	Pro	Leu	Leu	Gly	Ala	Ser	Leu	Asp	Ser	Ser	Gly	Gly	Lys	Lys
			530			535					540				
Leu	Leu	Gly	Leu	Ser	Tyr	Asp	Glu	Lys	His	Gln	Trp	Gln	Pro	Ile	Tyr
545					550					555					560
Gly	Ser	Thr	Pro	Val	Thr	Pro	Thr	Gly	Ser	Trp	Glu	Met	Gly	Lys	Arg
				565					570					575	
Tyr	His	Val	Val	Leu	Thr	Met	Ala	Asn	Lys	Ile	Gly	Ser	Val	Tyr	Ile

Asp	Gly	Glu	Pro	Leu	Glu	Gly	Ser	Gly	Gln	Thr	Val	Val	Pro	Asp	Gly
		595					600					605			
Arg	Thr	Pro	Asp	Ile	Ser	His	Phe	Tyr	Val	Gly	Gly	Tyr	Gly	Arg	Ser
	610					615					620				
Asp	Met	Pro	Thr	Ile	Ser	His	Val	Thr	Val	Asn	Val	Leu	Leu	Tyr	
	625				630					635				640	
Asn	Arg	Gln	Leu	Asn	Ala	Glu	Glu	Ile	Arg	Thr	Leu	Phe	Leu	Ser	Gln
				645					650					655	
Asp	Leu	Ile	Gly	Thr	Glu	Ala	His	Met	Gly	Ser	Ser	Ser	Gly	Ser	Ser
			660					665					670		
Ala	His	Ser	Thr	Pro	Ser	Thr	Pro	Ala	Asp	Asn	Gly	Ala	His	Ser	Thr
		675					680					685			
Pro	Ser	Thr	Pro	Ala	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Thr	Pro
	690					695					700				
Ala	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Gly	Asp	Asn	Gly
	705				710					715					720
Ala	His	Ser	Thr	Pro	Ser	Thr	Pro	Gly	Asp	Ser	Ser	Ala	His	Ser	Thr
				725					730					735	
Pro	Ser	Thr	Pro	Ala	Asp	Asn	Gly	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro
			740					745					750		
Ala	Asp	Ser	Asn	Ala	His	Ser	Thr	Pro	Ser	Thr	Pro	Ala	Asp	Asn	Gly
	755						760					765			
Ala	His	Ser	Thr	Pro	Ser	Thr	Pro	Ala	Asp	Asn	Gly	Ala	His	Ser	Thr
	770					775					780				
Pro	Ser	Thr	Pro	Gly	Asp	Asn	Gly	Ala	His	Ser	Thr	Pro	Ser	Thr	Pro
	785				790					795					800
Gly	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Thr	Pro	Ala	Asp	Asn	Gly
				805					810					815	
Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Ala	Asp	Ser	Asn	Ala	His	Ser	Thr
				820					825				830		
Pro	Ser	Thr	Pro	Gly	Asp	Asn	Gly	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro
			835				840					845			
Ala	Asp	Ser	Asn	Ala	His	Ser	Thr	Pro	Ser	Thr	Pro	Ala	Asp	Ser	Ser
	850					855					860				
Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Gly	Asp	Asn	Gly	Ala	His	Ser	Thr
	865				870					875					880
Pro	Ser	Ala	Pro	Ala	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro
				885					890					895	
Gly	Asp	Asn	Gly	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Ala	Asp	Asn	Gly
			900					905					910		
Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Gly	Asp	Ser	Asn	Ala	His	Ser	Thr
		915					920					925			
Pro	Ser	Thr	Pro	Ala	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Thr	Pro
		930				935						940			
Ala	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Gly	Asp	Asn	Gly
				945		950				955					960
Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Ala	Asp	Ser	Ser	Ala	His	Ser	Thr
				965					970					975	
Pro	Ser	Ile	Pro	Gly	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro
			980					985						990	
Ala	Asp	Ser	Ser	Ala	His	Ser	Thr	Pro	Ser	Ala	Pro	Gly	Asp	Asn	Gly
		995				1000						1005			
Ala	His	Ser	Thr	Pro	Ser	Thr	Pro	Ala	Asp	Asn	Gly	Ala	Asn	Gly	Thr
	1010					1015					1020				
Val	Leu	Ile	Leu	His	Asp	Gly	Ala	Ala	Phe	Ser	Ala	Phe	Ser	Gly	Gly
	1025				1030					1035					1040
Gly	Leu	Leu	Leu	Cys	Ala	Gly	Ala	Leu	Leu	Leu	His	Val	Phe	Val	Met
				1045				1050						1055	
Ala	Val	Phe	Phe												
			1060												

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1929 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGCTGGCAG CCGGATCGAG CCGAGTTGAG CTGTTTAAGC GGCAAAGCTC GAAGGTGCCA 60
TTTGAAGAGG ACGGCAAAGT CACCGAGCGG GTTGTCCACT CGTTCGCGCT CCCCGCCCTT 120
GTTAATGTGG ACGGGGTGAT GGTGCGCATC GCGGACGCTC GCTACGAAAC ATCCAATGAC 180
AACTCCCTCA TTGATACGGT GCGGAAGTAC AGCGTGGACG ATGGGGAGAC GTGGGAGACC 240
CAAATTGCCA TCAAGAACAG TCGTGCAATC TCTGTTTCTC GTGTGGTGGA TCCCACAGTG 300
ATTGTGAAGG GCAACAAGCT TTACGTCTCTG GTTGAAGCT ACAACAGTTC GAGGAGCTAC 360
TGGACGTCGC ATGGTGATGC GAGAGACTGG GATATCTCTG TTGCCGTTGG TGAGGTCAAG 420
AAGTCCACTG CGGGCGGCAA GATAACTGCG AGTATCAAAAT GGGGGAGCCC CGTGTCAGTG 480
AAGGAATTTT TTTCGGGCGA AATGGAAGGA ATGCACAAA ATCAATTTCT TGGCGGTGCA 540
GGTGTGTTGCA TTGTGGCGTC CAACGGGAAT CTTGTGTACC CTGTGCAGGT TACGAACAAA 600
AAGAAGCAAG TTTTTCCTCA GATCTTCTAC TCGGAAGACG AGGGCAAGAC GTGGAAGTTT 660
GGGAAGGGTA GGAGCGCTTT TGGCTGCTCT GAACCTGTGG CCCTTGAGTG GGAGGGGAAG 720
CTCATCATAA ACACGAGT TGAATATCGC CGCCGTCTGG TGTACGAGTC CAGTGACATG 780
GGGAATTCGT GGCTGGAGGC TGTCGGCAGC CTCTCACGTC TGTGGGGCCC CTCACCAAAA 840
TCGAACCAGC CCGGCAGTCA GAGCAGCTTC ACTGCCGTGA CCATCGAGGG AATGCGTGTT 900
ATGCTCTTCA CACACCCGCT GAATTTTAAG GGAAGGTGGC TGCGCGACCG ACTGAACCTC 960
TGGCTGACGG ATAACCAGCG CATTTATAAC GTTGGGCAAG TATCCATTGG TGATGAAAAT 1020
TCCGCTACA GCTCCGCTCT GTACAAGGAT GATAAGCTGT ACTGTTTGCA TGAGATCAAC 1080
AGTAACGAGG TGTACAGCCT TGTTTTTCGC CGCCTGGTTG GCGAGCTACG GATCATTAAA 1140
TCAGTGCTGC AGTCCTGGAA GAATTGGGAC AGCCACCTGT CCAGCATTTG CACCCCTGCT 1200
GATCCAGCCG CTTCGTCGTC AGAGCGTGGT TGTGGTCCCG CTGTCAACCAC GGTGTTGCTT 1260
GTTGGCTTTT TGTCGCACAG TGCCACCAAA ACCGAATGGG AGGATGCGTA CCGCTGCGTG 1320
AACGCAAGCA CCGCAAATGC GGAGAGGGTT CCGAACGGTT TGAAGTTTGC GGGGGTTGGC 1380
GGAGGGGCGC TTTGGCCGGT GAGCCAGCAG GGGCAGAATC AACGGTATCG CTTTGCAAAAC 1440
CACGCGTTCA CCGTGGTGGC GTCGGTGACG ATTCACGAGG TTCCGAGCGT CGCGAGTCTT 1500
TTGCTGGGTG CGAGCCTGGA CTCTTCTGGT GGCAAAAAAC TCCTGGGGCT CTCGTACGAC 1560
GAGAGGCACC AGTGGCAGCC AATATACGGA TCAACGCCGG TGACGCCGAC CGGATCGTGG 1620
GAGATGGGTA AGAGGTACCA CGTGGTTCTT ACGATGGCGA ATAAAAATTGG CTCGAGTAC 1680
ATTGATGGAG AACCTCTGGA GGGTTCAGGG CAGACCGTTG TGCCAGACGA GAGGACGCCT 1740
GACATCTCCC ACTTCTACGT TGGCGGGTAT AAAAGGAGTG ATATGCCAAC CATAAGCCAC 1800
GTGACGGTGA ATAATGTTCT TCTTTACAAC CGTCAGCTGA ATGCGAGGA GATCAGGACC 1860
TTGTTCTTGA GCCAGGACCT GATTGGCAGC GAAGCACACA TGGACAGCAG CAGCGACACG 1920
AGTGCCTGA 1929

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 642 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Leu Ala Pro Gly Ser Ser Arg Val Glu Leu Phe Lys Arg Gln Ser
1 5 10 15
Ser Lys Val Pro Phe Glu Lys Asp Gly Lys Val Thr Glu Arg Val Val
20 25 30
His Ser Phe Arg Leu Pro Ala Leu Val Asn Val Asp Gly Val Met Val
35 40 45
Ala Ile Ala Asp Ala Arg Tyr Glu Thr Ser Asn Asp Asn Ser Leu Ile
50 55 60
Asp Thr Val Ala Lys Tyr Ser Val Asp Asp Gly Glu Thr Trp Glu Thr
65 70 75 80
Gln Ile Ala Ile Lys Asn Ser Arg Ala Ser Ser Val Ser Arg Val Val
85 90 95
Asp Pro Thr Val Ile Val Lys Gly Asn Lys Leu Tyr Val Leu Val Gly
100 105 110

Ser	Tyr	Asn	Ser	Ser	Arg	Ser	Tyr	Trp	Thr	Ser	His	Gly	Asp	Ala	Arg
		115					120					125			
Asp	Trp	Asp	Ile	Leu	Leu	Ala	Val	Gly	Glu	Val	Thr	Lys	Ser	Thr	Ala
	130					135					140				
Gly	Gly	Lys	Ile	Thr	Ala	Ser	Ile	Lys	Trp	Gly	Ser	Pro	Val	Ser	Leu
	145				150					155					160
Lys	Glu	Phe	Phe	Pro	Ala	Glu	Met	Glu	Gly	Met	His	Thr	Asn	Gln	Phe
			165						170					175	
Leu	Gly	Gly	Ala	Gly	Val	Ala	Ile	Val	Ala	Ser	Asn	Gly	Asn	Leu	Val
			180					185					190		
Tyr	Pro	Val	Gln	Val	Thr	Asn	Lys	Lys	Gln	Val	Phe	Ser	Lys	Ile	
	195					200					205				
Phe	Tyr	Ser	Glu	Asp	Glu	Gly	Lys	Thr	Trp	Lys	Phe	Gly	Lys	Gly	Arg
	210					215					220				
Ser	Ala	Phe	Gly	Cys	Ser	Glu	Pro	Val	Ala	Leu	Glu	Trp	Glu	Gly	Lys
	225				230					235					240
Leu	Ile	Ile	Asn	Thr	Arg	Val	Asp	Tyr	Arg	Arg	Arg	Leu	Val	Tyr	Glu
			245						250					255	
Ser	Ser	Asp	Met	Gly	Asn	Ser	Trp	Leu	Glu	Ala	Val	Gly	Thr	Leu	Ser
			260					265					270		
Arg	Val	Trp	Gly	Pro	Ser	Pro	Lys	Ser	Asn	Gln	Pro	Gly	Ser	Gln	Ser
		275					280						285		
Ser	Phe	Thr	Ala	Val	Thr	Ile	Glu	Gly	Met	Arg	Val	Met	Leu	Phe	Thr
	290					295					300				
His	Pro	Leu	Asn	Phe	Lys	Gly	Arg	Trp	Leu	Arg	Asp	Arg	Leu	Asn	Leu
	305				310					315					320
Trp	Leu	Thr	Asp	Asn	Gln	Arg	Ile	Tyr	Asn	Val	Gly	Gln	Val	Ser	Ile
			325						330					335	
Gly	Asp	Glu	Asn	Ser	Ala	Tyr	Ser	Ser	Val	Leu	Tyr	Lys	Asp	Asp	Lys
			340					345					350		
Leu	Tyr	Cys	Leu	His	Glu	Ile	Asn	Ser	Asn	Glu	Val	Tyr	Ser	Leu	Val
		355					360					365			
Phe	Ala	Arg	Leu	Val	Gly	Glu	Leu	Arg	Ile	Ile	Lys	Ser	Val	Leu	Gln
	370					375					380				
Ser	Trp	Lys	Asn	Trp	Asp	Ser	His	Leu	Ser	Ser	Ile	Cys	Thr	Pro	Ala
	385				390					395					400
Asp	Pro	Ala	Ala	Ser	Ser	Ser	Glu	Arg	Gly	Cys	Gly	Pro	Ala	Val	Thr
			405						410					415	
Thr	Val	Gly	Leu	Val	Gly	Phe	Leu	Ser	His	Ser	Ala	Thr	Lys	Thr	Glu
			420					425					430		
Trp	Glu	Asp	Ala	Tyr	Arg	Cys	Val	Asn	Ala	Ser	Thr	Ala	Asn	Ala	Glu
		435					440					445			
Arg	Val	Pro	Asn	Gly	Leu	Lys	Phe	Ala	Gly	Val	Gly	Gly	Gly	Ala	Leu
	450					455					460				
Trp	Pro	Val	Ser	Gln	Gln	Gly	Gln	Asn	Gln	Arg	Tyr	Arg	Phe	Ala	Asn
	465				470					475					480
His	Ala	Phe	Thr	Val	Val	Ala	Ser	Val	Thr	Ile	His	Glu	Val	Pro	Ser
			485						490					495	
Val	Ala	Ser	Pro	Leu	Leu	Gly	Ala	Ser	Leu	Asp	Ser	Ser	Gly	Gly	Lys
			500					505					510		
Lys	Leu	Leu	Gly	Leu	Ser	Tyr	Asp	Glu	Arg	His	Gln	Trp	Gln	Pro	Ile
		515					520					525			
Tyr	Gly	Ser	Thr	Pro	Val	Thr	Pro	Thr	Gly	Ser	Trp	Glu	Met	Gly	Lys
	530					535					540				
Arg	Tyr	His	Val	Val	Leu	Thr	Met	Ala	Asn	Lys	Ile	Gly	Ser	Glu	Tyr
	545				550					555					560
Ile	Asp	Gly	Glu	Pro	Leu	Glu	Gly	Ser	Gly	Gln	Thr	Val	Val	Pro	Asp
			565						570					575	
Glu	Arg	Thr	Pro	Asp	Ile	Ser	His	Phe	Tyr	Val	Gly	Gly	Tyr	Lys	Arg
			580					585					590		
Ser	Asp	Met	Pro	Thr	Ile	Ser	His	Val	Thr	Val	Asn	Asn	Val	Leu	Leu
		595					600					605			
Tyr	Asn	Arg	Gln	Leu	Asn	Ala	Glu	Glu	Ile	Arg	Thr	Leu	Phe	Leu	Ser
	610					615					620				
Gln	Asp	Leu	Ile	Gly	Thr	Glu	Ala	His	Met	Asp	Ser	Ser	Ser	Asp	Thr
	625				630					635					640
Ser	Ala														

0905909 091001
T08T60 6065560

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTTCTAGAA TGCTGGCACC CGGATCGAGC

30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTGTGCGACA AAAAGCCAAC AAGACCAACC

30

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACTGAACCTC TGGCTGACGG ATAACCAGC

29

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTTCTCGAGT CAGGCACTCG TGTCGCTGCT

30

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGCAAGTAT CCATTGGTGA TGAAAATTCC GCCTACAGCT

40

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TACAGCTTAT CATCCTTGTA CAGGACGGAG CTGTAGGCGG

40